Pooling strength amongst limited datasets using hierarchical Bayesian analysis, with application to pyroclastic density current mobility metrics

Statistics in Volcanology
Volume 2
April 20, 2016

Pooling strength amongst limited datasets using hierarchical Bayesian analysis, with application to pyroclastic density current mobility metrics

Sarah E.Ogburn[1,2], James Berger[3], Eliza S. Calder[4], Danilo Lopes[5], Abani Patra[6], E. Bruce Pitman[7], Regis Rutarindwa[8], Elaine Spiller[8], Robert L. Wolpert[3]

[1]*State University of New York, University at Buffalo, Department of Geology, USA*
[2]*Now at: U.S. Geological Survey and U.S. Agency for International Development, Volcano Disaster Assistance Program (VDAP), Vancouver, WA, USA*
[3]*Duke University, Department of Statistical Sciences, Durham, NV, USA*
[4]*University of Edinburgh, School of Geosciences, UK*
[5]*Federal University of São Carlos, Brazil*
[6]*State University of New York, University at Buffalo, Department of Aerospace and Mechanical Engineering, USA*
[7]*State University of New York, University at Buffalo, Department of Mathematics, USA*
[8]*Marquette University, Department of Mathematicsm, Statistics, and Computer Science, Milwaukee, WI, USA*

Keywords: Augustine volcano, Colima volcano, Domehaz, Flowdat, hierarchical Bayesian analysis, mobility metrics, pyroclastic density currents, Semeru volcano, Soufrière Hills Volcano, Unzen volcano, volcanic hazards

Correspondence:
Sarah E . Ogburn: sogburn@usgs.gov
U.S. Geological Survey and U.S. Agency for International Development, Volcano Disaster Assistance Program (VDAP), Vancouver, WA 98683.

Citation:
Ogburn, S. E., Berger, J., Calder, E. S., Lopes, D., Patra, A., Pitman, E. B., Rutarindwa, R., Spiller, E., and Wolpert, R. L. (2016) Pooling strength amongst limited datasets using hierarchical Bayesian analysis, with application to pyroclastic density current mobility metrics, *Statistics in Volcanology* 2.1 : 1 – 26. DOI: http://dx.doi.org/10.5038/2163-338X.2.1

Abstract

In volcanology, the sparsity of datasets for individual volcanoes is an important problem, which, in many cases, compromises our ability to make robust judgments about future volcanic hazards. In this contribution we develop a method for using hierarchical Bayesian analysis of global datasets to combine information across different volcanoes and to thereby improve our knowledge at individual volcanoes. The method is applied to the assessment of mobility metrics for pyroclastic density currents in order to better constrain input parameters and their related uncertainties for forward modeling. Mitigation of risk associated with such flows depends upon accurate forecasting of possible inundation areas, often using empirical models that rely on mobility metrics measured from the deposits of past flows, or on the application of computational models, several of which take mobility metrics, either directly or indirectly, as input parameters. We use hierarchical Bayesian modeling to leverage the global record of mobility metrics from the FlowDat database, leading to considerable improvement in the assessment of flow mobility where the data for a particular volcano is sparse. We estimate the uncertainties involved and demonstrate how they are improved through this approach. The method has broad applicability across other areas of volcanology where relationships established from broader datasets can be used to better constrain more specific, sparser, datasets. Employing such methods allows us to use, rather than shy away from, limited datasets, and allows for transparency with regard to uncertainties, enabling more accountable decision-making.

Introduction

Efforts in quantitative volcanic hazards assessment (QVHA) are currently being bolstered by a number of ongoing initiatives to compile important databases such as the Global Volcanism Program database (*Global Volcanism Program*, 2013), WOVOdat (*Venezky & Newhall*, 2007), Geologic Survey of Japan (GSJ) Quaternary and Active volcanoes databases (*Geological Survey of Japan and the National Institute of Advanced Industrial Science and Technology (AIST)*, 2013), LaMEVE (*Crosweller et al.*, 2012), DomeHaz (*Ogburn et al.*, 2012, 2015), and FlowDat (*Ogburn*, 2012, 2014). These efforts collectively reflect a growing understanding of the value that is added by undertaking global analysis. Challenges remain, however, in dealing with variable data quality, sparse data for particular volcanic systems, and quantification of uncertainty. Some of these outstanding issues can be dealt with by exploring and developing statistical methods, which can not only improve our predictive capacity for future eruptions but can also contribute to advancing our scientific understanding of the volcanic processes involved.

Data sparsity, in particular, is a ubiquitous problem when assessing volcanic hazards (*Siebert et al.*, 2010). Indeed, *Siebert et al.* (2010) posit that *poorly known, thickly vegetated, long-quiescent volcanoes that have had no historical activity ... may be the most dangerous of all.* The record of activity at any given volcano may be incomplete or heavily biased due to inadequate or differential preservation and exposure of the deposits, or the history of nearby human settlement (*e.g., Crosweller et al.*, 2012; *Brown et al.*, 2014; *Kiyosugi et al.*, 2015; *Whelley et al.*, 2015). Other practical issues, such as accessibility and remoteness (*e.g., Whelley et al.*, 2015), also hinder investigation and therefore influence data completeness. In many cases, scientific interest in a given system is driven by significant observable volcanic activity, while small magnitude or effusive activity is often poorly recorded (*Deligne et al.*, 2010; *Furlan*, 2010; *Siebert et al.*, 2010; *Crosweller et al.*, 2012; *Brown et al.*, 2014). Often, newly active volcanoes, especially those that had previously been dormant (*e.g.,* Chaitén, Chile, in 2008; *Alfano et al.*, 2011; *Watt et al.*, 2013), may be poorly understood and may simply lack sufficient information on which to base assessments about renewed and future behavior.

The issues discussed above often result in information concerning a particular type of phenomenon (such as pyroclastic density currents) being plentiful at some well-studied volcanoes but very limited at others. Two

end-member approaches to deal with this problem are, *1)* to assume that particular phenomena have similar characteristics at every volcano, and thus use information from the global record of all volcanoes, or *2)* to assume that particular phenomena at different volcanoes behave dissimilarly, and use only the information from a given volcano. Often, however, it is reasonable to assume that a particular volcanic phenomenon, while not identical across volcanoes, is controlled by similar processes, and can be assumed to vary according to some probability distribution. This allows one to *borrow* information from a global database, leading to better quantification of uncertainty and improved accuracy in hazard assessment at a particular volcano. The statistical methodology for doing this is hierarchical Bayesian analysis (*Allenby et al.*, 2005). Bayesian approaches to volcanic hazard assessment have been used successfully for event tree construction (*Marzocchi et al.*, 2008, 2010) and have recently been expanded using hierarchical Bayesian methods (*Sheldrake*, 2014).

In this work, we use hierarchical Bayesian methods to augment statistical analysis of the mobility of pyroclastic density currents. Specifically for this work, our interest is in dense, concentrated dome-collapse pyroclastic density currents. Pyroclastic density currents (PDCs) are hot avalanches of rock and gas which, due to their ability to travel great distances at high speeds, are among the most destructive volcanic hazards. This effort, in part, is motivated by the need for more robust characterization of the mobility relationships of PDCs for different volcanic systems. Mitigation of risk associated with these phenomena depends upon accurate forecasting of possible flow paths and inundation areas, often using empirical models that rely on mobility metrics (*e.g.,* the energy cone model, *Malin & Sheridan* (1982); PFz, *Widiwijayanti et al.* (2008)) or the application of computational flow models (*e.g.,* TITAN2D *Patra et al.* (2005); VolcFlow, *Kelfoun & Druitt* (2005)). Linear regression of mobility metrics such as the Heim coefficient (height dropped/runout length of a PDC, or H/L) or the relationship between the area inundated by a PDC and its volume, often informs such models, sometimes as direct model inputs (*e.g.,* the energy cone model, PFz), or indirectly as proxies for input parameters (*e.g.,* basal friction angle in TITAN2D, constant resisting shear stress in VolcFlow). There are many examples where such data has been used successfully to simulate and replicate the behavior of past events (*Kelfoun & Druitt*, 2005; *Widiwijayanti et al.*, 2008; *Charbonnier & Gertisser*, 2009, 2012; *Murcia et al.*, 2010; *Sheridan et al.*, 2010; *Capra et al.*, 2011; *Ogburn*, 2014). However, the use of such data as input parameters in forward modeling of future hazards is compromised by the relative dearth of information on large volume events and the scarcity of data from remote, under-studied, or recently active volcanoes. When eruptive activity initiates at a newly active volcano, for which little PDC data is available, forward modeling by simply substituting PDC mobility parameters from other volcanoes is of tenuous merit, as local source conditions and topographic effects influence flow mobility (*Stinton*, 2014; *Charbonnier & Gertisser*, 2011; *Lube et al.*, 2011; *Ogburn*, 2014) and inundation estimates will have high uncertainties. Instead, what is required are more accountable approaches to enable the use of the limited existing data to their maximum potential while also quantifying the associated uncertainty.

We develop a method using hierarchical Bayesian analysis to leverage the global record of mobility metrics from the FlowDat mass flow database (*Ogburn*, 2012, 2014). Strength is borrowed from the global record to understand mobility characteristics at specific volcanoes, leading to considerable improvement in assessments where data for a particular volcano is sparse. First, the background to the problem of assessing mobility of PDCs and how PDC mobility metrics are used with, and subsequently propagated through, flow modeling, is presented in Section 1. The hierarchical Bayesian analysis of the compiled data is presented in Section 2, and the results are discussed in Section 3. The variables and abbreviations used throughout are presented in Appendix A and a detailed step-by-step methodology is provided in Appendix B.

1 Mobility metrics for mass flows

1.1 Frictional vs. resisting shear stress models

The most widely used mobility metric for concentrated mass flows of *(e.g.,* volcanic and non-volcanic debris avalanches, dome- and column-collapse PDCs) is the Heim coefficient (*Heim*, 1932), commonly denoted as (H/L), where H is the vertical fall height traversed by a flow and L is the runout length. (H/L) is equivalent to the coefficient of friction following a Mohr-Coulomb friction model, in which shear stress at the initiation of failure is proportional to the normal stress.

According to Mohr-Coulomb friction models, the mass or volume, V, of the flow should be irrelevant to mobility, and the coefficient of friction should be a function of material properties. Numerous studies of real deposits, however, have shown a linear inverse relationship between $\log(V)$ of a mass flow (of any type) and $\log(H/L)$ (*Heim*, 1932; *Scheller*, 1971; *Scheidegger*, 1973; *Hsü*, 1975), with large volume flows demonstrably being more mobile than small volume flows.

An alternative to the frictional model approach is the class of the constant resisting shear stress models. In these models, the mobility of mass flows is described by a constant resisting shear stress (CRS), or yield strength, and the planimetric area, A_p, is related to $V^{2/3}$ via scaling arguments (*Hungr*, 1990; *Iverson et al.*, 1998; *Dade & Huppert*, 1998; *Calder et al.*, 1999). This model indicates a relationship between inundated area and resisting shear stress, suggesting a yield stress rheology (*Kilburn & Sørenson*, 1998; *Crosta et al.*, 2003; *Griswold & Iverson*, 2008).

Both of these metrics (H/L and A_p vs. $V^{2/3}$) have been applied to PDC mobility with success (*Sparks*, 1976; *Nairn & Self*, 1978; *Francis & Baker*, 1977; *Sheridan*, 1979; *Begét & Limke*, 1988; *Fisher & Schmincke*, 1984; *Hayashi & Self*, 1992; *Calder et al.*, 1999; *Cole et al.*, 2002; *Vallance et al.*, 2010; *Charbonnier & Gertisser*, 2011) and have become standard mobility metrics with which to compare and contrast PDC behavior, especially, but not exclusively, those of concentrated PDCs.

1.2 Mobility metrics for flow modeling

Many empirical flow inundation models are based directly on measurements of (H/L) or (A_p vs. $V^{2/3}$). *Hsü* (1975), *Sheridan* (1979), and *Malin & Sheridan* (1982) first used the *energy-line* or *energy-cone* concept (which is defined by H/L). This concept has been applied at a variety of volcanoes (*e.g., Sheridan & Malin*, 1983; *Wadge & Isaacs*, 1988; *Höskuldsson & Cantagrel*, 1994; *Alberico et al.*, 2002; *Sheridan et al.*, 2004) and also forms the basis for the FLOW2D and FLOW3D computer models (*e.g., Kover & Sheridan*, 1993; *Martin del Pozzo et al.*, 1995; *Sheridan & Macías*, 1995; *Hooper & Mattioli*, 2001) which base shear resistance on basal friction (taken directly from H/L), viscosity, and turbulence.

(H/L) also informs computational flow models that use a Coulomb friction law, including TITAN2D (*Patra et al.*, 2005), which have built upon the work of *Savage & Hutter* (1989), who used Coulomb friction laws in conjunction with depth-averaged equations for mass and momentum. The Heim coefficient can therefore provide a guideline for choosing appropriate basal friction input angles for different flow volumes for TITAN2D (*Ogburn*, 2008, 2014; *Charbonnier & Gertisser*, 2012; *Charbonnier et al.*, 2015).

LAHARZ and PFZ use semi-empirical equations for planimetric area ($A_p = cV^{2/3}$) and cross-sectional area ($A_{xs} = CV^{2/3}$) to predict lahar (*Iverson et al.*, 1998), debris flow, rock avalanche (*Griswold & Iverson*, 2008) and PDC (*Widiwijayanti et al.*, 2008) inundation using empirically derived coefficients (c and C) from a variety of mass flow deposits worldwide. These relationships also form the basis of flow models using constant shear stress instead of constant friction (*e.g.,* VolcFlow, *Kelfoun & Druitt*, 2005).

With increasing application of these respective flow modeling approaches, it is now timely and appropriate to undertake more considered approaches to understanding and quantifying the uncertainty related to the use of mobility metrics as model inputs. This work has been driven by our specific interest in constraining the basal

friction input parameter required by TITAN2D when undertaking ensemble runs for generating probabilistic hazards maps (*Bayarri et al.*, 2009; *Spiller et al.*, 2014; *Bayarri et al.*, 2015), by using the (H/L)-volume mobility relationships for block-and-ash flows from the FlowDat database. The application of the method developed can, however, be applied widely.

2 Statistical analyses

Herein, we present a method using hierarchical Bayes modeling to leverage the global record of mobility metrics for PDCs, which can aid in cases where data for a particular volcano is sparse. We use the FlowDat database of mass flow mobility metrics (*Ogburn*, 2012, 2014), which is current through 2014. From FlowDat (*Ogburn*, 2012), 4 volcanoes were selected with plentiful (H/L) data, planimetric areas, and volume data for dome-collapse PDCs (14 to 80 flows): *(i)* Colima Volcano, Mexico (data from: *Saucedo et al.*, 2002, 2004, 2010), *(ii)* Merapi Volcano, Indonesia (data from: *Boudon et al.*, 1993; *Bourdier & Abdurachman*, 2001; *Schwarzkopf et al.*, 2005; *Charbonnier & Gertisser*, 2011; *Charbonnier et al.*, 2013; *Komorowski et al.*, 2013), *(iii)* Soufrière Hills Volcano, Montserrat (data from: *Calder et al.*, 1999; *Cole et al.*, 2002; *Hards et al.*, 2008; *Komorowski et al.*, 2010; *Loughlin et al.*, 2010; *Cole et al.*, 2014), and *(iv)* Unzen Volcano, Japan (data from: *Nakada et al.*, 1999; *Takarada*, 2008) (Figure 1). Volcanoes with sparse data were also used: *(i)* for the (H/L) plot, Semeru Volcano, Indonesia, (data from *Thouret et al.*, 2007), and *(ii)* for the (A_p vs. $V^{2/3}$) plot, Augustine Volcano, Alaska (data from: *Kamata et al.*, 1991; *Vallance et al.*, 2010; *Global Volcanism Program*, 2013) and Unzen Volcano, Japan, (data from *Nakada et al.*, 1999). These flows are all dense, concentrated dome-collapse PDCs (block and ash flows), for which it is reasonable to assume broadly similar flow behavior. Error was rarely reported by the sources of the data, but is shown as error bars where available. However, the error bars were often smaller than the markers themselves.

For the frictional model of mobility (H/L vs. V), the strong linear relationship between the logarithm of PDC volume and the logarithm of the coefficient of friction suggests the use of a linear model, such as a regression model

$$y = \alpha + \beta x + \epsilon, \qquad \epsilon \overset{iid}{\sim} N(0, \sigma^2)$$

where x is the log-volume[1], y is the log-coefficient of friction (H/L), α and β are the intercept and slope of the regression line, and ϵ is random error. Graphically, this model corresponds to fitting a straight line through all of the data $\boldsymbol{y}$ in Figure 1, which minimizes the errors between estimated and observed values. This approach corresponds to one end-member option, that is, to assume that the relationship between the coefficient of friction and flow volume for block-and-ash flows is constant at every volcano, and thus use information from all the volcanoes to fit a regression.

Alternatively, one could fit separate regression lines for each of the J volcanoes, namely

$$y_j = \alpha_j + \beta_j x + \epsilon, \qquad \epsilon \overset{iid}{\sim} N(0, \sigma_j^2)\,,$$

based on the data $\boldsymbol{y}_j$ from volcano j alone. The result of separate regression fits is shown in Figure 1. This approach represents the alternative end-member option, that is, to assume that the relationship between the coefficient of friction and volume at different volcanoes is unrelated, and thus uses only the information from a given volcano to fit a regression.

Likewise, to fit the constant resisting shear stress relationship (A_p vs. $V^{2/3}$), we apply the same models to the transformed volume $(V^{2/3})$ and planimetric area data (A_p) by letting x be the $\log(V^{2/3})$ and y be the $\log(A_p)$. The analysis in the next section is described in terms of the frictional relationship, but applied in

[1] Actually $x = \log_{10}(\text{volume}/10^{5.5})$. This x-origin then corresponds volume of $10^{5.5}\text{m}^3$, roughly where the slope and intercept are least correlated.

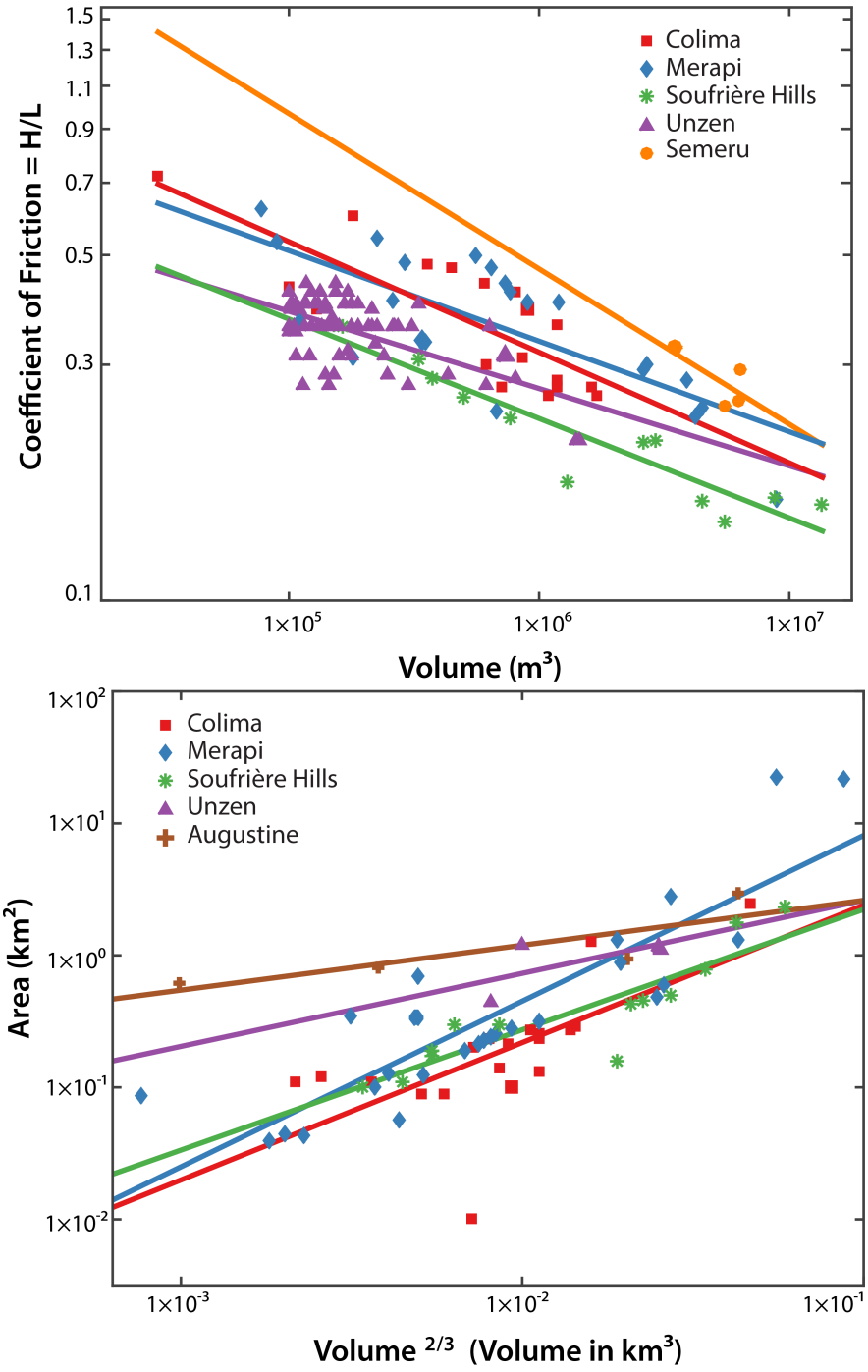

Figure 1: Data from all volcanoes considered for each of the two respective relationships along with their respective linear regression lines. Upper plot shows coefficient of friction (H/L) vs. volume (V). Colima, Merapi, Soufrière Hills, and Unzen have plentiful data, while data for Semeru is sparse. Lower plot shows planimetric area (A_p) vs. transformed volume $(V^{2/3})$. Colima, Merapi, and Soufrière Hills have plentiful data, while data for Unzen and Semeru is sparse. Error bars on all values are smaller than the markers themselves and errors for volumes were only reported for Soufrière Hills. Note that not all PDCs had both (H/L) and (A_p) values reported in the literature.

an identical manner for the constant resisting shear stress relationship using the appropriate definitions for x and y. Furthermore, the hierarchical analysis presented in the next section could prove useful for *any* linear relationship suggested by transformations of volcanic datasets; the frictional and constant resisting shear stress relationships for dome-collapse PDCs used here are just two pertinent examples.

2.1 Hierarchical Bayesian model

In situations where it is unclear whether to fit an overall regression model or separate regression models, it has become common statistical practice to use the hierarchical or multilevel approach, which is a happy medium between these end-member alternatives. Hierarchical modeling is carried out via Bayesian analysis, wherein a *prior* probability distribution is chosen to describe knowledge about the unknown model parameters (here the various regression parameters); this distribution will then be updated by the data to form *posterior* probability distributions of the unknown model parameters.

The version of hierarchical modeling utilized here links together the separate regressions by assuming that the regression line slopes arose from the common normal distribution (part of the prior distribution)

$$\beta_j \quad \text{are i.i.d.} \quad N(\mu, \tau^2)\,,$$

with unknown hyper-mean (the mean of the prior distribution) μ and hyper-variance (the variance of the prior distribution) τ^2. Note that, if $\tau^2 = 0$, then all the β_j would be equal, so we would be back to the case of a single regression. At the other extreme, as $\tau^2 \to \infty$, this model would yield the same answers as the separate regression models. The performance of the hierarchical model, in situations such as this, is typically better than that of either of the two extremes.

An initial presumption is that little is known about μ and τ^2 (a vague prior distribution will be used for these parameters), but more will be learned about them from the data through their posterior distribution and they, in turn, will affect the posterior distribution of the β_j.

If data were plentiful at each volcano, there would be little need (but also no harm) in employing the hierarchical model, as the effect of the posterior distribution of μ and τ^2 on the β_j would then be minimal. When data is sparse for one or more volcanoes, however, the gains with the hierarchical approach can be considerable. For example, from the top panel of Figure 1 it can be seen that there are only four data points from Semeru for a very narrow range of PDC volumes, and attempting to fit a separate regression to just four points will lead to a very uncertain result. In contrast, the hierarchical modeling approach allows for *borrowing strength* from the other volcanoes in estimating Semeru's regression line slope (because of the assumption that all slopes arose from a common normal distribution), and will be seen to result in much tighter credible intervals for the regression line for Semeru.

To complete the specification of the hierarchical model, *prior* distributions for the other unknown parameters in the model need to be chosen. Whereas the regression coefficients from Figure 1 appear quite related, the intercepts, α_j, seem considerably more variable. A hierarchical model for the intercepts could be utilized, but since there would be little gain, an objective constant prior distribution $\pi^O(\alpha_1, \ldots, \alpha_J) = 1$ is employed instead; although this objective prior does not induce any sharing of intercept information across volcanoes, the changes in the slope parameters through their hierarchical modeling will influence the intercepts.

In developing prior distributions for the regression variances σ_j^2, it is important to consider that the PDC data represented in Figure 1 come from both highly channelized and unchannelized (unconfined) flows; both flows experience different frictional forces and exhibit different mobilities (*Ogburn*, 2014; *Charbonnier & Gertisser*, 2011; *Stinton*, 2014). Modeling by *Stinton* (2014) using TITAN2D showed that flows confined in synthetic channels had longer runouts, higher velocities, and shorter travel times than flows simulated over synthetic unconfined terrain. *Lube et al.* (2011) found a similar topographic effect on the (A_p vs. $V^{2/3}$) metric, whereby

increasing the proportion of the flow which escaped from a channel strongly increased this ratio. This was explained by the order of magnitude difference between the thickness of channel-confined and unconfined portions of the deposits. *Lube et al.* (2011) and *Charbonnier & Gertisser* (2011) also noted a change in mobility metrics as flows inundating the same drainage progressively filled and reduced the carrying capacity of the channel, resulting in higher proportions of unconfined deposits. The degree of channelization of particular PDCs is not trivial to determine in a quantitative sense, as PDCs often exhibit a combination of both channelized and unchannelized transport that varies downstream. Additionally, many of the traditional metrics (*i.e.*, plan aspect ratio) can be heavily influenced by the width of, and thus confinement imposed by, the channels themselves (*Ogburn*, 2014). However, both qualitative descriptions of PDCs from the literature and transect measurements of channelization for a limited number of PDCs indicate that both the Merapi and Colima datasets contain PDCs with lower degrees of channelization than the datasets from the other three volcanoes. PDCs at Merapi and Colima also tend to inundate multiple channels, while those elsewhere typically travel down a single channel.

These differences are also apparent in the data. Indeed, Table 1 gives the results of separate regressions at the five volcanoes, and the mean square residuals (MSR) are very similar for the three volcanoes with dominantly channelized flows and are much smaller than the MSR for the volcanoes with dominantly unchannelized flow deposits. The higher MSR for unchannelized flows or those that inundate multiple channels makes intuitive sense, as these flows travel over extremely varied topography, with greater variation in slope and surface roughness than flows which travel down channels. An exception are PDCs at Augustine, which were mainly unchannelized, but each PDC was emplaced over relatively similar substrates of snow and ice, reflected in the low MSR for those flows. We have, therefore, grouped these PDCs with the channelized flows. It would be natural to have a separate variance for the channelized and the unchannelized flow data. Thus, we assign Merapi and Colima a common variance σ_C^2 and the other volcanoes common variance σ_U^2, with the two variances being unknown.

The equivalent slope, intercept, error information for the (A_p vs. $V^{2/3}$) relationship is summarized in Table 2. For this analysis, we also apply the channelized/unchannelized grouping to specify σ_C^2 and σ_U^2.

Table 1: Linear regression parameters and MSR for each volcano for (H/L vs. V) relationship.

Volcano	**Lin. Reg. Slope**	**Lin. Reg. Intercept**	**MSR** ($\times 10^{-4}$)
Colima*	-0.224	-0.386	66.5
Merapi*	-0.183	-0.384	95.2
Soufrière Hills	-0.201	-0.531	24.8
Unzen	-0.156	-0.493	26.3
Semeru	-0.314	-0.172	24.3

* indicates volcanoes with flows which are generally **unchannelized**, otherwise flows are channelized

To complete the Bayesian model, prior distributions are needed for σ_C^2 and σ_U^2 and for the hyperparameters μ and τ^2 from the hierarchical prior. For these parameters we utilize a standard objective prior, the *reference prior*, $\pi^R(\mu, \sigma_\beta^2, \sigma_C^2, \sigma_U^2)$; this is given in Appendix B. The reference prior is chosen so as to minimize the influence of the prior distribution on the analysis, *i.e.*, to ensure that the posterior distribution of the model parameters only reflects what the data has to say.

This completes the specification of the Bayesian hierarchical model, and one now simply applies Bayes

Table 2: Linear regression parameters and MSR for each volcano for the (A_p vs. $V^{2/3}$) relationship.

Volcano	Lin. Reg. Slope	Lin. Reg. Intercept	MSR
Colima*	1.041	1.421	0.142
Merapi*	1.256	2.165	0.128
Soufrière Hills	0.912	1.260	0.042
Unzen	0.553	0.971	0.076
Augustine*	0.340	0.757	0.039

* indicates volcanoes with flows which are generally **unchannelized**, otherwise flows are channelized, with the exception of Augustine PDCs (see text)

theorem to obtain the posterior distribution of all unknowns parameters, given all the data $\boldsymbol{y}$, as

$$\begin{aligned}\pi(\alpha_1,\ldots,\alpha_J,\beta_1,\ldots,\beta_J,\mu,\tau^2,\sigma_C^2,\sigma_U^2 \mid \boldsymbol{y}) \propto \prod_{j=1}^{J} f(\boldsymbol{y}_j \mid \alpha_j,\beta_j,\sigma_j^2) \\ \times\pi^O(\alpha_1,\ldots,\alpha_J)\pi^R(\mu,\sigma_\beta^2,\sigma_C^2,\sigma_U^2)\prod_{j=1}^{J} N(\beta_j \mid \mu,\tau^2)\,,\end{aligned} \tag{1}$$

where $f(\boldsymbol{y}_j \mid \alpha_j,\beta_j,\sigma_j^2)$ is the likelihood arising from the data at volcano j and the σ_j^2 are either the channelized or unchannelized variance.

2.2 Analysis

There are no closed form analytical expressions for estimates of unknown parameters or for credible intervals, but there is a relatively straightforward Markov Chain Monte Carlo (MCMC) method, described in Appendix B, for drawing samples from the posterior distribution in (1). From this set of samples,

$$\{(\alpha_1^i,\ldots,\alpha_J^i,\beta_1^i,\ldots,\beta_J^i,\mu^i,(\tau^2)^i,(\sigma_C^2)^i,(\sigma_U^2)^i), \quad i=1,\ldots,m\}\,,$$

all desired inferences can be performed.

The typical parameter estimate would be the posterior mean, computed as the average of all of the samples; enough samples are typically chosen ($m = 10^6$ was used in the computations herein) that the numerical error in this computation is negligible. Similarly a 95% credible interval, for example, would be formed as the interval containing the central 95% of the ordered sample. Even more informatively, the entire posterior distribution of a parameter could be approximated by simply making a histogram of the sample values. These histograms are illustrated in Appendix B (Figure 5 and Figure 6). Note, in particular, from Figure 6 that the channelized and unchannelized variances do seem to be quite different.

3 Geophysical results and discussion

The relationship between coefficient of friction and volume can be studied in several ways from the posterior sample of parameters. For volcano j, we have a sample $\{(\alpha_j^i,\beta_j^i),\ i=1,\ldots,m\}$ of the intercepts and slopes. This yields a sample from the posterior distribution of all regressions lines, illustrated in Figure 2.

Samples of regression lines are useful for the computation of inundation probabilities from PDCs; for example, where it is necessary to consider different possible mobilities for flows over a range of volumes. Samples

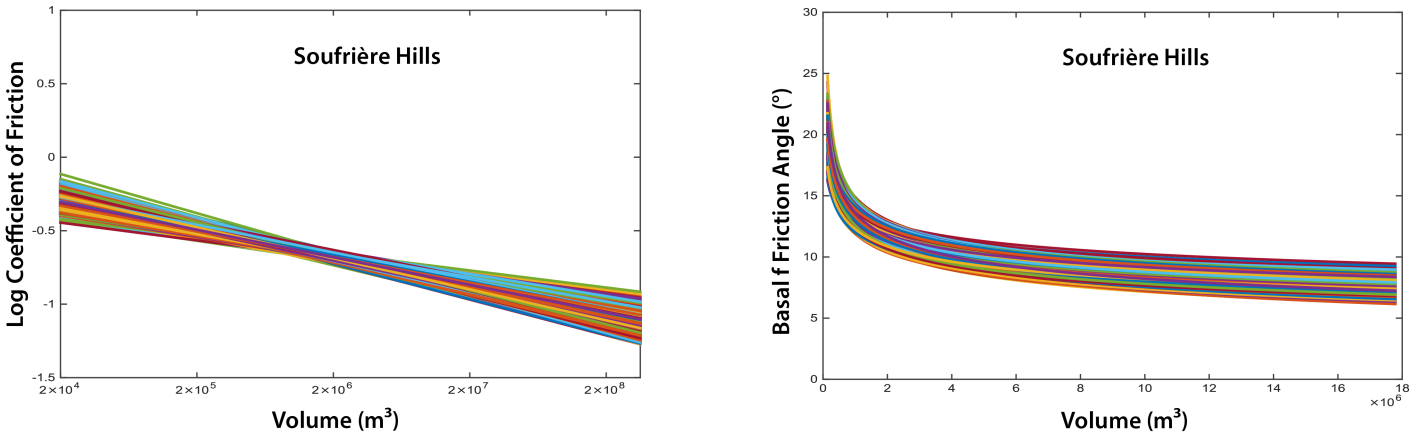

Figure 2: Both figures represent samples from the hierarchical linear regression model of the frictional relationship applied to the data, but show the same sample curves on different scales. Left plot shows coefficient of friction and volume each on a log scale (which the linear model was fit to). Right plot shows Basal friction angle (calculated as arctan of the coefficient of friction) versus volume on a linear scale (1×10^6 m^3).

from regression lines can be used directly for empirical models such as the energy line/cone method or for estimating the basal friction input parameter for a geophysical model like TITAN2D (*Bayarri et al.*, 2009; *Ogburn*, 2014; *Spiller et al.*, 2014) or the constant resisting shear stress input parameter in VolcFlow (*Ogburn*, 2014). Furthermore, using regression samples generated by this method allows one to account for uncertainty in probabilistic assessments of PDC inundation.

Figure 3 gives, for each volcano, a posterior summary consisting of the regression line corresponding to the posterior median values of the sample regression lines (the solid red line); this would be the natural estimated regression line from the Bayesian analysis. 95% credible intervals (the dashed red lines) are also shown and are obtained, at each volume value V, by taking the central 95% interval of values of $\alpha_j^i + \beta_j^i \log_{10}(V)$, over the posterior samples.

For comparison, the confidence intervals on the regression function from classical individual regressions are also given in Figure 3, with the solid black line being the standard estimated regression function and the dashed black lines being the standard 95% confidence intervals. As expected, for the volcanoes with abundant data, there is not much difference between the hierarchical model regression summaries and the classical regressions. But, for Semeru, which had only four data points all of which are closely clustered in volume, the differences found would affect the results of a probabilistic analyses, with the hierarchical approach providing tighter uncertainty estimates. This conclusion is, of course, predicated on the scientific judgment that the slope of the Semeru regression line is related to the slopes of the others, but this is reasonable.

For the (A_p vs. $V^{2/3}$) relationship, again we summarize the posterior distribution of the hierarchical linear model. Figure 4 gives, for each volcano, the regression line corresponding to the posterior median values of the sample regression lines (the solid red line), and 95% credible intervals (the dashed red lines) formed, at each volume value V, by taking the central 95% interval of values of $\alpha_j^i + \beta_j^i \log_{10}(V)$, over the posterior samples. And again, for comparison, the confidence intervals on the regression function from classical individual regressions are also given, with the solid black line being the standard estimated regression function and the dashed black lines being the standard 95% confidence intervals.

Again, for Figure 4 we have two volcanoes with limited data, Unzen (three data points) and Augustine (four data points). The reduction in uncertainty obtained through the hierarchical linear model is rather different for the two cases. Although the 95% credible intervals from the hierarchical model are reduced in both cases

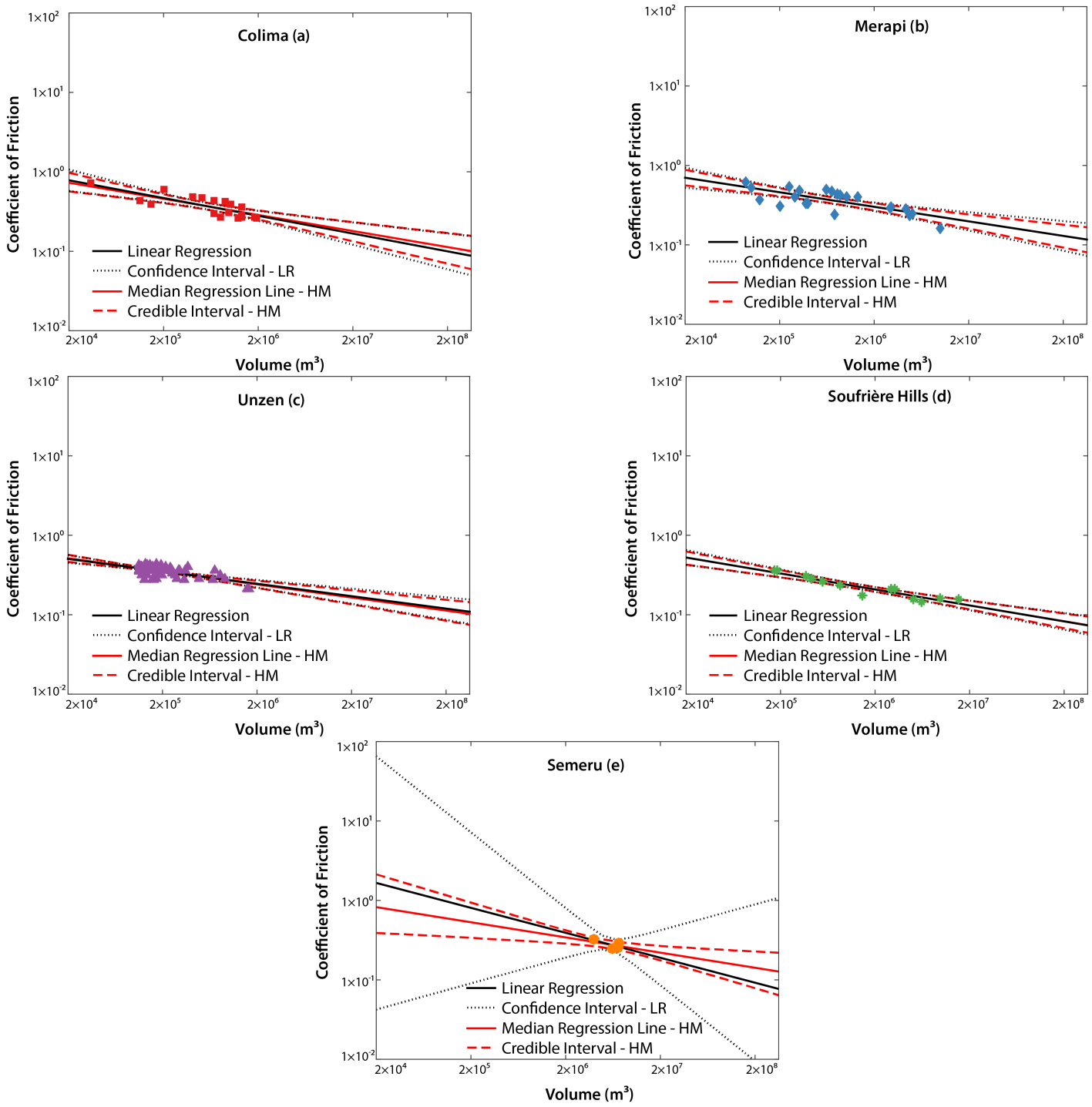

Figure 3: Comparison of the 95% confidence intervals (black dotted line) on the regression line for each individual volcano (black solid lines) and credible intervals (red dotted line) obtained from the hierarchical model (red solid line) as applied to the coefficient of friction vs. volume relationship. PDCs in (a) and (b) were considered unchannelized; PDCs in (c) and (d) were considered channelized in this analysis. PDCs from Semeru (e) were also considered channelized, but with only four data points.

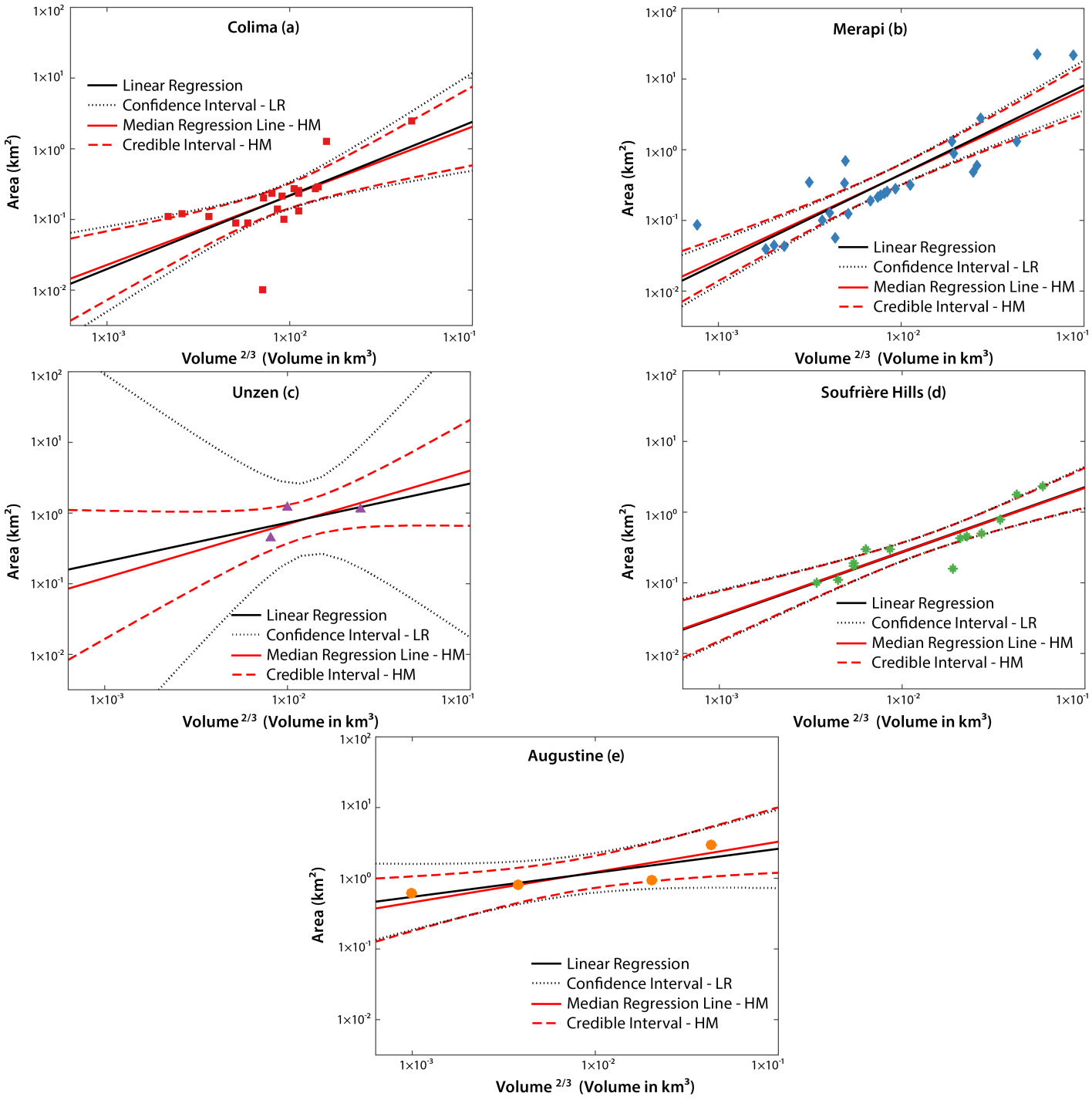

Figure 4: Comparison of the 95% confidence intervals (black dotted line) on the regression line for each individual volcano (black solid lines) and credible intervals (red dotted line) obtained from the hierarchical model (red solid line) as applied to the (A_p) vs. ($V^{2/3}$). PDCs in (a) and (b) were considered unchannelized; PDCs in (c) and (d) were considered channelized in this analysis. Augustine (e) produced unchannelized flows which traveled over surfaces of snow and ice.

(as well as for Colima), the improvements are much more dramatic for Unzen, which has data points that are tightly clustered in volume. This is a case, much like Semeru in the frictional relationship, where *borrowing strength* from other volcanoes via the hierarchical analysis greatly reduces uncertainty in fitting an inferential relationship.

This type of approach is broadly applicable to other types of mass flows (debris avalanches, lahars, or column-collapse PDCs, for example) or other types of data entirely (ash-dispersion metrics, for example), but it is important that the datasets selected describe phenomena that are similar. This work focused only on dense, dome-collapse PDCs which are considered to have broadly similar emplacement dynamics; and accounted for dissimilarities (*i.e.*, differences in channelization) by allowing for different variances. However, the more similar the phenomena at different volcanoes, the better the method is able to reduce uncertainty. The selection of appropriate data is thus subject to scientific judgment.

Finally, it is important to note that this work does not seek to recommend one mobility metric over another, but rather to illustrate the usefulness of the hierarchical Bayesian approach for different types of commonly reported mobility metrics that inform model inputs. The choice of which mobility metric, conceptual model, or computational model is most appropriate for different types of mass flows is a matter of much debate (*e.g., Dade & Huppert*, 1998; *Kilburn & Sørenson*, 1998; *Legros*, 2002; *Kelfoun & Druitt*, 2005) and detailed comparisons of these models can be found elsewhere in the literature (*Kelfoun & Druitt*, 2005; *Charbonnier & Gertisser*, 2012; *Ogburn*, 2014). It is also worth noting here that for larger volume and more dilute flows, fluidization and turbulence plays a more dominant role and that the mobility metrics and modeling tools referred to here are of limited utility.

Conclusions

Understanding the past behavior of a particular volcano is the foundation upon which assessments of potential future hazards are based. However, complete and robust datasets are very rare, and really only exist for a handful of very well-studied volcanoes. Additionally, newly active volcanoes may produce hazards with poorly constrained characteristics. This problem can be handled by, *1)* using only data from a particular volcano (which may be sparse, and thus introduce large uncertainties into hazard assessments), or *2)* using the global record of volcanoes (which may ignore or downplay any particularities of the volcano in question). The hierarchical Bayesian method for analyzing mobility metrics presented herein allows one to achieve a happy medium between these two approaches by not only using data from a particular volcano, but also by *borrowing strength* from the global record of PDC behavior and thus greatly reducing the uncertainty for volcanoes with sparse data.

Acknowledgments

The authors would like to thank two anonymous reviewers and editor Charles Connor for their helpful reviews and feedback. This work was supported in part by the following grants: NSF EAR-0809543 NSF NERC RiftVolc NE/L013932/1 and NSF EAR-1331353

References

Alberico, Ines, Lirer, Lucio, Petrosino, Paola & Scandone, Roberto (2002) A methodology for the evaluation of long-term volcanic risk from pyroclastic flows in Campi Flegrei (Italy). *Journal of Volcanology and Geothermal Research*, **116**(1–2):63–78. URL http://dx.doi.org/10.1016/S0377-0273(02)00211-1. 4

Alfano, Fabrizio, Bonadonna, Costanza, Volentik, Alain C. M., Connor, Charles B., Watt, Sebastian F. L., Pyle, David M. & Connor, Laura J. (2011) Tephra stratigraphy and eruptive volume

of the May, 2008, Chaitén eruption, Chile. *Bulletin of Volcanology*, **73**(5):613–630. URL http://dx.doi.org/10.1007/s00445-010-0428-x. 2

Allenby, Greg M., Rossi, Peter E. & McCulloch, Robert E. (2005) Hierarchical Bayes models: A practitioners guide. URL http://dx.doi.org/10.2139/ssrn.655541. Available at Social Science Research Network (SSRN). 3

Bayarri, M.J., Berger, J.O., Calder, E.S., Dalbey, K., Lunagomez, S., Patra, A.K., Pitman, E.B., Spiller, E.T. & Wolpert, R.L. (2009) Using statistical and computer models to quantify volcanic hazards. *Technometrics*, **51**(4):402–413. URL http://dx.doi.org/10.1198/TECH.2009.08018. 5, 10

Bayarri, M. J., Berger, J. O., Calder, E. S., Patra, Abani K., Pitman, E Bruce, Spiller, E. T. & Wolpert, Robert L. (2015) Probabilistic quantification of hazards: A methodology using small ensembles of physics-based simulations and statistical surrogates. *International Journal for Uncertainty Quantification*, **5**(4):297–325. URL http://dx.doi.org/10.1615/Int.J.UncertaintyQuantification.2015011451. 5

Begét, James E. & Limke, Anthony J. (1988) Two-dimensional kinematic and rheological modeling of the 1912 pyroclastic flow, Katmai, Alaska. *Bulletin of Volcanology*, **50**(3):148–160. URL http://dx.doi.org/10.1007/BF01079679. 4

Berger, James O. & Bernardo, José M. (1992) Reference priors in a variance components problem. *In* P.K. Goel & N.S. Iyengar (eds.) *Bayesian Analysis in Statistics and Econometrics*, volume 75 of *Lecture Notes in Statistics*, 177–194. Springer, New York, NY. ISBN 978-0-387-97863-5. URL http://dx.doi.org/10.1007/978-1-4612-2944-5_10. 22

Boudon, Georges, Camus, Guy, Gourgaud, Alain & Lajoie, Jean (1993) The 1984 nuée-ardente deposits of Merapi volcano, Central Java, Indonesia: stratigraphy, textural characteristics, and transport mechanisms. *Bulletin of Volcanology*, **55**(5):327–342. URL http://dx.doi.org/10.1007/BF00301144. 5

Bourdier, J.-L. & Abdurachman, E.K. (2001) Decoupling of small-volume pyroclastic flows and related hazards at Merapi volcano, Indonesia. *Bulletin of Volcanology*, **63**(5):309–325. URL http://dx.doi.org/10.1007/s004450100133. 5

Brown, Sarah Krystyna, Crosweller, Helen Sian, Sparks, Robert Stephen John, Cottrell, Elizabeth, Deligne, Natalia Irma, Guerrero, Natalie Ortiz, Hobbs, Laura, Kiyosugi, Koji, Loughlin, Susan Clare, Siebert, Lee & Takarada, Shinji (2014) Characterisation of the Quaternary eruption record: analysis of the Large Magnitude Explosive Volcanic Eruptions (laMEVE) database. *Journal of Applied Volcanology*, **3**:22. URL http://dx.doi.org/10.1186/2191-5040-3-5. 2

Calder, Eliza S., Cole, Paul D., Dade, W. Brian, Druitt, Timothy H., Hoblitt, Richard P., Huppert, H. E., Ritchie, Lucie J., Sparks, R. Stephen J. & Young, Simon R. (1999) Mobility of pyroclastic flows and surges at the Soufrière Hills Volcano, Montserrat. *Geophysical Research Letters*, **26**(5):537–540. URL http://dx.doi.org/10.1029/1999GL900051. 4, 5

Capra, L., Manea, V. C., Manea, M. & Norini, G. (2011) The importance of digital elevation model resolution on granular flow simulations: a test case for Colima volcano using TITAN2D computational routine. *Natural Hazards*, **59**(2):665–680. URL http://dx.doi.org/10.1007/s11069-011-9788-6. 3

Casella, George & George, Edward I. (1992) Explaining the Gibbs sampler. *The American Statistician*, **46**(3):167–174. URL http://dx.doi.org/10.1080/00031305.1992.10475878. 22

Charbonnier, Sylvain J., Germa, Aurelie, Connor, Chuck B., Gertisser, Ralf, Komorowski, Jean-Christophe, Preece, Katie, Lavigne, Frank, Dixon, Tim & Connor, Laura (2013) Evaluation of the impact of the 2010 pyroclastic density currents at Merapi volcano from high-resolution satellite imagery, field investigation and numerical simulations. *Journal of Volcanology and Geothermal Research*, **261**:295–315. URL http://dx.doi.org/10.1016/j.jvolgeores.2012.12.021. 5

Charbonnier, S. J. & Gertisser, R. (2009) Numerical simulations of block-and-ash flows using the Titan2D flow model: examples from the 2006 eruption of Merapi Volcano, Java, Indonesia. *Bulletin of Volcanology*, **71**(8):953–959. URL http://dx.doi.org//10.1007/s00445-009-0299-1. 3

Charbonnier, Sylvain J. & Gertisser, Ralf (2011) Deposit architecture and dynamics of the 2006 block-and-ash flows of Merapi Volcano, Java, Indonesia. *Sedimentology*, **58**(6):1573–1612. URL http://dx.doi.org/10.1111/j.1365-3091.2011.01226.x. 3, 4, 5, 7, 8

Charbonnier, Sylvain J. & Gertisser, Ralf (2012) Evaluation of geophysical mass flow models using the 2006 block-and-ash flows of Merapi Volcano, Java, Indonesia: Towards a short-term hazard assessment tool. *Journal of Volcanology and Geothermal Research*, **231–232**:87–108. URL http://dx.doi.org/10.1016/j.jvolgeores.2012.02.015. 3, 4, 13

Charbonnier, S. J., Palma, J. L. & Ogburn, S. (2015) Application of 'shallow-water' numerical models for hazard assessment of volcanic flows: the TITAN2D model. *Revista Geológica de América Central*, 107–128. URL http://dx.doi.org/10.15517/rgac.v0i52.19021. 4

Cole, P.D., Smith, P.J., Stinton, A.J., Odbert, H.M., Bernstein, M.L., Komorowski, J.-C. & Stewart, R. (2014) Vulcanian explosions at Soufrière Hills Volcano, Montserrat between 2008 and 2010. *In* G. Wadge, R. Robertson & B. Voight (eds.) *The Eruption of Soufrière Hills Volcano, Montserrat, from 2000 to 2010*, volume 39, 93–111. Geological Society of London, Memoirs. URL http://dx.doi.org/10.1144/M39.5. 5

Cole, P. D., Calder, E. S., Sparks, R. S. J., Clarke, A. B., Druitt, T. H., Young, S. R., Herd, R. A., Harford, C. L. & Norton, G. E. (2002) Deposits from dome-collapse and fountain-collapse pyroclastic flows at Soufrière Hills Volcano, Montserrat. *In* T.H. Druitt & B.P. Kokelaar (eds.) *The eruption of Soufrière Hills Volcano, Montserrat, from 1995–1999*, volume 21, 231–262. Geological Society of London, Memoirs. URL http://dx.doi.org/10.1144/GSL.MEM.2002.021.01.11. 4, 5

Crosta, Giovanni B., Cucchiaro, Samuel & Frattini, Paulo (2003) Validation of semi-empirical relationships for the definition of debris-flow behavior in granular materials. *In* D. Rickenmann & C.L. Chen (eds.) *Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment*, 821–831. Millpress, Rotterdam, NL. ISBN 978-90-77017-78-4. URL http://www.millpress.nl/shop/abooks/DHFM/pdf/E2_04.pdf. 4

Crosweller, H.S., Arora, B., Brown, S.K., Cottrell, E., Deligne, N.I., Guerrero, N.O., Hobbs, L., Kiyosugi, K., Loughlin, S.C., Lowndes, J., Nayembil, M., Siebert, L., Sparks, R.S.J., Takarada, S. & Venzke, E. (2012) Global database on large magnitude explosive volcanic eruptions (LaMEVE). *Journal of Applied Volcanology*, **1**(4):13. URL http://dx.doi.org/10.1186/2191-5040-1-4. 2

Dade, W. Brian & Huppert, Herbert E. (1998) Long-runout rockfalls. *Geology*, **26**(9):803–806. URL http://dx.doi.org/10.1130/0091-7613(1998)026<0803:LRR>2.3.CO;2. 4, 13

Deligne, N. I., Coles, S. G. & Sparks, R. S. J. (2010) Recurrence rates of large explosive volcanic eruptions. *Journal of Geophysical Research*, **115**(B6). URL http://dx.doi.org/10.1029/2009JB006554. 2

Fisher, Richard V. & Schmincke, Hans-Ulrich (1984) *Pyroclasic Rocks*. Springer-Verlag, Berlin Heidelberg. ISBN 978-3-540-51341-4. URL http://dx.doi.org/10.1007/978-3-642-74864-6. 4

Francis, P. W. & Baker, M. C. W. (1977) Mobility of pyroclastic flows. *Nature*, **270**:164–165. URL http://dx.doi.org/10.1038/270164a0. 4

Furlan, Claudia (2010) Extreme value methods for modelling historical series of large volcanic magnitudes. *Statistical Modelling*, **10**(2):113–132. URL http://dx.doi.org/10.1177/1471082X0801000201. 2

Geological Survey of Japan and the National Institute of Advanced Industrial Science and Technology (AIST) (2013) Catalog of eruptive events during the last 10,000 years in Japan, version 2.2. URL https://translate.google.com/translate?hl=en&sl=ja&u=https://gbank.gsj.jp/volcano/eruption/&prev=search. 2

Global Volcanism Program (2013) Volcanoes of the World (VOTW), v. 4.4.3, E. Venzke (ed.). URL http://dx.doi.org/10.5479/si.GVP.VOTW4-2013. Downloaded 30 Mar 2016. 2, 5

Griswold, Julia P. & Iverson, Richard M. (2008) Mobility and statistics and automated hazard mapping for debris flows and rock avalanches, version 1.1. Scientific Investigations Report 2007-5276, U. S. Geological Survey. URL http://pubs.usgs.gov/sir/2007/5276/. 4

Hards, V., Strutt, M., De Angelis, S., Ryan, G., Christopher, T., Syers, T. & Bass, V. (2008) Report to the Scientific Advisory Committee on activity at Soufrière Hills Volcano Montserrat: April 2008. *Montserrat Volcano Observatory Open File Report*, **OFR 09-01**. 5

Hayashi, J. N. & Self, S. (1992) A comparison of pyroclastic flow and debris avalanche mobility. *Journal of Geophysical Research B: Solid Earth*, **97**(B6):9063–9071. URL http://dx.doi.org/10.1029/92JB00173. 4

Heim, Albert (1932) Bergsturz und Menschenleben. *Zürich Vierteljahrsschrift*, **77**:218. URL http://www.gra-nat.ch/system/media/2410/original/77BB20_0.pdf?1422022711. [English translation by N. A. Skermer, 1989. Landslides and human lives: Vancouver, B.C., BiTech Publishers, 195p]. 4

Hooper, Donald M. & Mattioli, Glen S. (2001) Kinematic modeling of pyroclastic flows produced by gravitational dome collapse at Soufrière Hills Volcano, Montserrat. *Natural Hazards*, **23**(1):65–86. URL http://dx.doi.org/10.1023/A:1008130605558. 4

Höskuldsson, Armann & Cantagrel, Jean Marie (1994) Volcanic hazards in the surroundings of Pico de Orizaba, eastern Mexico. *Natural Hazards*, **10**(3):197–219. URL http://dx.doi.org/10.1007/BF00596142. 4

Hsü, Kenneth J. (1975) Catastrophic debris streams (Sturtzstroms) generated by rockfalls. *Geological Society of America Bulletin*, **86**(1):129–140. URL http://dx.doi.org/10.1130/0016-7606(1975)86<129:CDSSGB>2.0.CO;2. 4

Hungr, Oldrich (1990) Mobility of rock avalanches. Report 46, National Research Institute for Earth Science and Disaster Prevention (NIED). URL http://dil-opac.bosai.go.jp/publication/nied_report/PDF/46/46hungr.pdf. 4

Iverson, Richard M., Schilling, Steven P. & Vallance, James W. (1998) Objective delineation of lahar-inundation hazard zones. *Geological Society of America Bulletin*, **110**(8):972–984. URL http://dx.doi.org/10.1130/0016-7606(1998)110<0972:ODOLIH>2.3.CO;2. 4

Kamata, H., Johnston, D.A. & Waitt, R.B. (1991) Stratigraphy, chronology, and character of the 1976 pyroclastic eruption of Augustine volcano, Alaska. *Bulletin of Volcanology*, **53**(6):407–419. URL http://dx.doi.org/10.1007/BF00258182. 5

Kelfoun, Karim & Druitt, Timothy H. (2005) Numerical modelling of the emplacement of the 7500 BP Socompa rock avalanche, Chile. *Journal of Geophysical Research B: Solid Earth*, **110**(B12):B12202. URL http://dx.doi.org/10.1029/2005JB003758. 3, 4, 13

Kilburn, Christopher R. J. & Sørenson, Søren-Aksel (1998) Runout lengths of sturztroms; the control of initial conditions and of fragment dynamics. *Journal of Geophysical Research B: Solid Earth*, **103**(B8):17877–17884. URL http://dx.doi.org/10.1029/98JB01074. 4, 13

Kiyosugi, Koji, Connor, Charles, Sparks, Robert Stephen John, Crosweller, Helen Sian, Brown, Sarah Krystyna, Siebert, Lee, Wang, Ting & Takarada, Shinji (2015) How many explosive eruptions are missing from the geologic record? Analysis of the quaternary record of large magnitude explosive eruptions in Japan. *Journal of Applied Volcanology*, **4**(1):17. URL http://dx.doi.org/10.1186/s13617-015-0035-9. 2

Komorowski, J.-C., Jenkins, S., Baxter, P.J., Picquout, A., Lavigne, F., Charbonnier, S.J., Gertisser, R., Preece, K., Cholik, N., Budi-Santoso, A. & Surono (2013) Paroxysmal dome explosion during the Merapi 2010 eruption: Processes and facies relationships of associated high-energy pyroclastic density currents. *Journal of Volcanology and Geothermal Research*, **261**:260–294. URL http://dx.doi.org/10.1016/j.jvolgeores.2013.01.007. 5

Komorowski, J.-C., Legendre, Y., Christopher, T., Bernstein, M., Stewart, R., Joseph, E., Fournier, N., Chardot, L., Finizola, A., Wadge, G., Syers, R., Williams, C. & Bass, V.A. (2010) Insights into processes and deposits of hazardous vulcanian explosions at Soufrière Hills Volcano during 2008 and 2009 (Montserrat, West Indies). *Geophysical Research Letters*, **37**(19):1–6. URL http://dx.doi.org/10.1029/2010GL042558. 5

Kover, Thomas P. & Sheridan, Michael F. (1993) Numerical models for pyroclastic flows of Mt. Unzen, Japan. *Geological Society of America Abstracts With Programs*, **25**(6):A268. 4

Legros, François (2002) The mobility of long-runout landslides. *Engineering Geology*, **63**(3–4):301–331. URL http://dx.doi.org/10.1016/S0013-7952(01)00090-4. 13

Loughlin, S.C., Luckett, R. & Ryan, G.A. (2010) An overview of lava dome evolution, dome collapse and cyclicity at Soufrière Hills Volcano, Montserrat, 2005–2007. *Geophysical Research Letters*, **37**(19):L00E16. URL http://dx.doi.org/10.1029/2010GL042547. 5

Lube, Gert, Cronin, Shane J. & Thouret, Jean-Claude (2011) Kinematic characteristics of pyroclastic density currents at Merapi and controls on their avulsion from natural and engineered channels. *Geological Society of America Bulletin*, **123**(5-6):1127–1140. URL http://dx.doi.org/10.1130/B30244.1. 3, 7, 8

Malin, Michael C. & Sheridan, Michael F. (1982) Computer-assisted mapping of pyroclastic surges. *Science*, **217**(4560):637–640. URL http://dx.doi.org/10.1126/science.217.4560.637. 3, 4

Martin del Pozzo, Ana Lillian, Sheridan, Michael, Barrera, David, Lugo Hubp, José & Vázquez Selem, Lorenzo (1995) Potential hazards from Colima volcano, Mexico. *Geofísica Internacional*, **34**(4):363–376. 4

Marzocchi, Warner, Sandri, Laura & Selva, Jacopo (2008) BET_EF: a probabilistic tool for long- and short-term eruption forecasting. *Bulletin of Volcanology*, **70**(5):623–632. URL http://dx.doi.org/10.1007/s00445-007-0157-y. 3

Marzocchi, Warner, Sandri, Laura & Selva, Jacopo (2010) BET_VH: a probabilistic tool for long-term volcanic hazard assessment. *Bulletin of Volcanology*, **72**(6):705–716. URL http://dx.doi.org/10.1007/s00445-010-0357-8. 3

Mengersen, Kerrie L., Robert, Christian P. & Guihenneuc-Jouyaux, Chantal (1999) MCMC convergence diagnostics: A review. *In* J.M. Bernardo, J.O. Berger, A.P. Dawid & A.F.M. Smith (eds.) *Bayesian Statistics 6, Proceedings of the Sixth Valencia International Meeting*, 415–440. Oxford University Press, Oxford. ISBN 9780198504856. 24

Murcia, H.F., Sheridan, M.F., Macías, J.L. & Cortés, G.P. (2010) TITAN2D simulations of pyroclastic flows at Cerro Machín Volcano, Colombia: Hazard implications. *Journal of South American Earth Sciences*, **29**(2):161–170. URL http://dx.doi.org/10.1016/j.jsames.2009.09.005. 3

Nairn, Ian A. & Self, Steve (1978) Explosive eruptions and pyroclastic avalanches from Ngauruhoe in February 1975. *Journal of Volcanology and Geothermal Research*, **3**(1–2):39–60. URL http://dx.doi.org/10.1016/0377-0273(78)90003-3. 4

Nakada, S., Shimizu, H. & Ohta, K. (1999) Overview of the 1990–1995 eruption at Unzen Volcano. *Journal of Volcanology and Geothermal Research*, **89**(1):1–22. URL http://dx.doi.org/10.1016/S0377-0273(98)00118-8. 5

Ogburn, S.E., Loughlin, S.C. & Calder, E.S. (2015) The association of lava dome growth with major explosive activity (VEI $\geq$ 4): DomeHaz, a global dataset. *Bulletin of Volcanology*, **77**:1–17. URL http://dx.doi.org/10.1007/s00445-015-0919-x. 2

Ogburn, Sarah E. (2008) *Potential hazards at Soufrière Hills Volcano, Montserrat: Northwards-directed dome-collapses and major explosive eruptions.* Master's thesis, State University of New York at Buffalo, Buffalo, NY. 4

Ogburn, Sarah E. (2012) Flowdat: Mass flow database. On Vhub.org at https://vhub.org/groups/massflowdatabase. 2, 3, 5

Ogburn, Sarah E. (2014) *Reconciling field observations of pyroclastic density currents with conceptual and computational analogs using a GIS and a newly developed global database.* Ph.D. thesis, State University of New York at Buffalo, Buffalo, NY. 2, 3, 4, 5, 7, 8, 10, 13

Ogburn, S. E., Loughlin, S.C. & Calder, E. S. (2012) Domehaz: Dome-forming eruptions database. On Vhub.org at https://vhub.org/groups/domedatabase. 2

Patra, A. K., Bauer, A. C., Nichita, C. C., Pitman, E. B., Sheridan, M. F. & Bursik, M. I. (2005) Parallel adaptive numerical simulation of dry avalanches over natural terrain. *Journal of Volcanology and Geothermal Research*, **139**(1–2):1–21. URL http://dx.doi.org/10.1016/j.jvolgeores.2004.06.014. 3, 4

Saucedo, R., Macías, J.L & Bursik, M.I. (2004) Pyroclastic flow deposits of the 1991 eruption of Volcán de Colima, Mexico. *Bulletin of Volcanology*, **66**(4):291–306. URL http://dx.doi.org/10.1007/s00445-003-0311-0. 5

Saucedo, R., Macías, J.L., Bursik, M.I., Mora, J.C., Gavilanes, J.C. & Cortes, A. (2002) Emplacement of pyroclastic flows during the 1998-1999 eruption of Volcán de Colima, México. *Journal of Volcanology and Geothermal Research*, **117**:129–153. URL http://dx.doi.org/10.1016/S0377-0273(02)00241-X. 5

Saucedo, R., Macías, J.L., Gavilanes, J.C., Arce, J.L, Komorowski, J.-C., Gardner, J.E. & Valdez-Moreno, G. (2010) Eyewitness, stratigraphy, chemistry, and eruptive dynamics of the 1913 Plinian eruption of Volcán de Colima, México. *Journal of Volcanology and Geothermal Research*, **191**(3-4):149–166. URL http://dx.doi.org/10.1016/j.jvolgeores.2010.01.011. 5

Savage, Stuart B. & Hutter, Kolumban (1989) The motion of a finite mass of granular material down a rough incline. *Journal of Fluid Mechanics*, **199**:177–215. URL http://dx.doi.org/10.1017/S0022112089000340. 4

Scheidegger, Adrian E. (1973) On the prediction of the release and velocity of catastrophic rockfalls. *Rock Mechanics*, **5**(4):231–236. URL http://dx.doi.org/10.1016/S0377-0273(02)00241-X. 4

Scheller, Erwin (1971) Beitrag zum Bewegungsverhalten grosser Berstürze. *Eclogae Geologicae Helvetiae*, **64**:195–202. URL http://www.e-periodica.ch/digbib/view?pid=egh-001:1971:64#258. 4

Schwarzkopf, L.M., Schmincke, H.-U. & Cronin, S.J. (2005) A conceptual model for block-and-ash flow basal avalanche transport and deposition, based on deposit architecture of 1998 and 1994 Merapi flows. *Journal of Volcanology and Geothermal Research*, **139**(1-2):117–134. URL http://dx.doi.org/10.1016/j.jvolgeores.2004.06.012. 5

Sheldrake, Tom (2014) Long-term forecasting of eruption hazards: A hierarchical approach to merge analogous eruptive histories. *Journal of Volcanology and Geothermal Research*, **286**:15–23. URL http://dx.doi.org/10.1016/j.jvolgeores.2014.08.021. 3

Sheridan, Michael F. (1979) Emplacement of pyroclastic flows: A review. *In* C.E. Chapin & W.E. Elston (eds.) *Ash-Flow Tuffs*, 125–136. Geological Society of America, Special Paper 180, Boulder, CO. URL http://dx.doi.org/10.1130/SPE180-p125. 4

Sheridan, Michael F., Hubbard, Bernard, Carrasco-Núñez, Gerardo & Siebe, Claus (2004) Pyroclastic flow hazard at Volcán Citlaltépetl. *Natural Hazards*, **33**(2):209–221. URL http://dx.doi.org/10.1023/B:NHAZ.0000037028.89829.d1. 4

Sheridan, Michael F. & Macías, José Luis (1995) Estimation of risk probability for gravity-driven pyroclastic flows at Volcán Colima, México. *Journal of Volcanology and Geothermal Research*, **66**(1–4):251–256. URL http://dx.doi.org/10.1016/0377-0273(94)00058-O. 4

Sheridan, Michael F. & Malin, Michael C. (1983) Application of computer-assisted mapping to volcanic hazard evaluation of surge eruptions: Vulcano, Lipari, and Vesuvius. *Journal of Volcanology and Geothermal Research*, **17**(1–4):187–202. URL http://dx.doi.org/10.1016/0377-0273(83)90067-7. 4

Sheridan, Michael F., Patra, Abani K., Dalbey, Keith & Hubbard, Bernard (2010) Probabilistic digital hazard maps for avalanches and massive pyroclastic flows using TITAN2D. *In GSA Special Papers*, volume 464 of *Stratigraphy and Geology of Volcanic Areas*, 281–291. Geological Society of America. URL http://dx.doi.org/10.1130/2010.2464(14). 3

Siebert, L., Simkin, T. & Kimberly, P. (2010) *Volcanoes of the World*. Smithsonian Institution, Third edition. URL http://www.ucpress.edu/ebook.php?isbn=9780520947931. [Digital edition available from California University Press]. 2

Sparks, R. Stephen J. (1976) Grain size variations in ignimbrites and implications for the transport of pyroclastic flows. *Sedimentology*, **23**(2):147–188. URL http://dx.doi.org/10.1111/j.1365-3091.1976.tb00045.x. 4

Spiller, E.T., Bayarri, M.J., Berger, J.O., Calder., E.S., Patra, A.K., Pitman, E.B. & Wolpert, R.L. (2014) Automating emulator construction for geophysical hazard maps. *SIAM/ASA Journal on Uncertainty Quantification*, **2**(1):126–152. URL http://dx.doi.org/10.1137/120899285. 5, 10

Stinton, Adam J. (2014) *Effects of Changes in Valley Geomorphology on the Behavior of Volcanic Mass-flows*. Ph.D. thesis, State University of New York at Buffalo, Buffalo, NY. 3, 7

Takarada, S. (2008) Personal communication. 5

Thouret, J.-C., Lavigne, F., Suwa, H., Sukatja, B. & Surono (2007) Volcanic hazards at Mount Semeru, East Java (Indonesia), with emphasis on lahars. *Bulletin of Volcanology*, **70**(2):221–244. URL http://dx.doi.org/10.1007/s00445-007-0133-6. 5

Vallance, James W., Bull, K. F. & Coombs, Michelle L. (2010) Pyroclastic flows, lahars and mixed avalanches generated during the 2006 eruption of Augustine Volcano: chapter 10. *In* J.A. Power, M.L. Coombs & J.T. Freymueller (eds.) *The 2006 eruption of Augustine Volcano, Alaska*, USGS Professional Paper 1769, 219–267. U.S. Geological Survey. URL http://pubs.usgs.gov/pp/1769/. 4, 5

Venezky, D.Y. & Newhall, C.G. (2007) WOVOdat design document: the schema, table descriptions, and create table statements for the database of worldwide volcanic unrest (WOVOdat Version 1.0). Open-File Report 2007-1117, U.S. Geological Survey. URL http://pubs.usgs.gov/of/2007/1117/. 2

Wadge, Geoff & Isaacs, Michael C. (1988) Mapping the volcanic hazards from the Soufrière Hills Volcano, Montserrat, West Indies, using an image processor. *Journal of the Geological Society*, **145**(4):541–551. URL http://dx.doi.org/10.1144/gsjgs.145.4.0541. 4

Watt, Sebastian F.L., Pyle, David M. & Mather, Tamsin A. (2013) Evidence of mid- to late-Holocene explosive rhyolitic eruptions from Chaitén Volcano, Chile. *Andean Geology*, **40**(2):216–226. URL http://dx.doi.org/10.5027/andgeoV40n2-a02. 2

Whelley, P. L., Newhall, C. G. & Bradley, K. E. (2015) The frequency of explosive volcanic eruptions in Southeast Asia. *Bulletin of Volcanology*, **77**(1):1–11. URL http://dx.doi.org/10.1007/s00445-014-0893-8. 2

Widiwijayanti, C., Voight, B., Hidayat, D. & Schilling, S. P. (2008) Objective rapid delineation of areas at risk from block-and-ash pyroclastic flows and surges. *Bulletin of Volcanology*, **71**(6):687–703. URL http://dx.doi.org/10.1007/s00445-008-0254-6. 3, 4

Appendix A

Table 3: Variables and abbreviations.

	Frictional model
V	volume of PDC
H	height dropped/vertical travel distance of PDC
L	runout length of PDC
H/L	Heim coefficient, coefficient of friction in the friction model
	Constant resisting shear stress model
A_p	Planimetric area of PDC inundation
	Statistical model
y	dependent variable: log-coefficient of friction (H/L), or planimetric area
x	independent variable: log-volume or $V^{2/3}$
α	intercept of the regression line
β	slope of the regression line
ϵ	random error
iid	is independent and identically distributed
$\sim$	has the distribution
$N(0, \sigma^2)$	a normal distribution with a mean of 0 and a variance σ^2
J	each of the J volcanoes
μ	hyper-mean, the mean of the prior distribution
τ^2	hyper-variance, the variance of the prior distribution
$\pi^O(\alpha_a, ..., \alpha_J) = 1$	objective constant prior distribution
σ^2_C	common variance for channelized PDCs
σ^2_U	common variance for unchannelized PDCs
$\pi^R(\mu, \sigma^2_\beta, \sigma^2_C, \sigma^2_U)$	reference prior
MSR	Mean square residual
MCMC	Markov Chain Monte Carlo

Appendix B

The technical details of the hierarchical Bayesian analysis are given herein. First, some notation: write the design matrix for the j^{th} regression (*i.e.*, the intercept constant 1 and the transformed volume input values) as (with n_j being the number of observations for Volcano j)

$$\boldsymbol{X}_j = \begin{pmatrix} 1 & x_{j1} \\ 1 & x_{j2} \\ \vdots & \vdots \\ 1 & x_{jn_j} \end{pmatrix},$$

and define (recalling that the σ_j^2 are σ_C^2 or σ_U^2 for the channelized and unchannelized volcanoes)

$$\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji},\ S_j = \sum_{i=1}^{n_j}(x_{ji}-\bar{x}_j)^2,\ \lambda_j = \frac{\tau^2}{\sigma_j^2},\ v_j = v_j(\sigma_j^2,\tau^2) = d_j + \tau^2,\ d_j = \frac{\sigma_j^2}{S_j},$$
$$\boldsymbol{v} = (v_1,\ldots,v_J),\ n = \sum_{j=1}^{J} n_j,\ \begin{pmatrix}\hat{\alpha}_j \\ \hat{\beta}_j\end{pmatrix} = (\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j'\boldsymbol{y}_j,\ \hat{\mu}(\boldsymbol{v}) = \frac{\sum_{j=1}^{J}\hat{\beta}_{j2}/v_j}{\sum_{j=1}^{J}1/v_j}.$$

The objective reference prior for the parameters $(\mu, \tau^2, \sigma_U^2, \sigma_C^2)$ is (*Berger & Bernardo*, 1992)

$$\pi(\mu,\tau^2,\sigma_U^2,\sigma_C^2) = \left(\frac{1}{\sigma_U^2\sigma_C^2}\right)\sqrt{\sum_{j=1}^{J}\frac{1}{v_j^2(\sigma_j^2,\tau^2)}}.$$

Then a Gibbs sampler (*Casella & George*, 1992) can be constructed as follows, to draw samples from the posterior distribution in (1).

Step 1. Draw the β_j, given σ_j^2, μ and τ^2, from the following distribution:

$$N\left(\hat{\beta}_j - \frac{(\hat{\beta}_j-\mu)}{1+\lambda_j S_j},\quad \frac{\sigma_j^2\lambda_j}{1+\lambda_j S_j}\right).$$

This is the marginal posterior distribution of β_j, given σ_j^2, μ and τ^2 (*i.e.*, α_j has been integrated out). Note that we could have also integrated out μ, but that should not be necessary because below we generate μ from its marginal posterior distribution with the βs integrated out.

Step 2. Draw the α_j, given σ_j^2 and β_j, from the $N\left(\hat{\alpha}_j - \bar{x}_j(\beta_j - \hat{\beta}_j),\ \ \sigma_j^2/n_j\right)$ distribution. This is the conditional posterior distribution of α_j, given σ_j^2 and β_j. (It happens to not depend on τ^2 or μ.)

Step 3a. Propose a value of σ_U^2, given the $\{\beta_j\}, j = 1, 2$, by drawing a random variable from the inverse gamma distribution with shape parameter $\alpha_U = (n_1 + n_2)/2$ and rate parameter

$$\beta_U = \frac{1}{2}\sum_{j=1}^{2}(y_{ji} - [\alpha_j + x_{ji}\beta_j])^2.$$

Draw a uniform random variable U on $(0, 1)$ and accept the proposed σ_U^2 if

$$U < \frac{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2,\tau^2)}}{\sqrt{\sum_{j=1}^{2} 1/v_j^2(0,\tau^2) + \sum_{j=3}^{5} 1/v_j^2(\sigma_j^2,\tau^2)}};$$

else discard σ_U^2 and propose a new σ_U^2, repeating as necessary until a σ_U^2 is accepted. This arises from the standard accept-reject algorithm because the numerator above, which is the unnormalized ratio of the target posterior distribution and the inverse gamma proposal distribution, is maximized at $\sigma_U^2 = 0$.

Step 3b. Propose a value of σ_C^2, given the $\{\beta_j\}, j = 3, 4, 5$, by drawing a random variable from the inverse gamma distribution with shape parameter $\alpha_C = (n_3 + n_4 + n_5)/2$ and rate parameter

$$\beta_C = \frac{1}{2}\sum_{j=3}^{5}(y_{ji} - [\alpha_j + x_{ji}\beta_j])^2 .$$

Draw a uniform random variable U on $(0, 1)$ and accept σ_C^2 if

$$U < \frac{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2, \tau^2)}}{\sqrt{\sum_{j=1}^{2} 1/v_j^2(\sigma_j^2, \tau^2) + \sum_{j=3}^{5} 1/v_j^2(0, \tau^2)}} ;$$

else discard σ_C^2 and draw a new σ_C^2, repeating as necessary until a σ_C^2 is accepted. The rationale is as in Step 3A. These steps yield draws from the conditional posterior distributions of σ_U^2 and σ_C^2, given the $\{\alpha_j, \beta_j\}$, and do not depend on the other parameters.

Step 4. Draw μ, given the σ_j^2 and τ^2, from the following distribution:

$$N\left(\hat{\mu}(\boldsymbol{v}), \quad \frac{1}{\sum_{j=1}^{J} 1/v_j}\right) .$$

This is the marginal posterior distribution of μ, given the σ_j^2 and τ^2, *i.e.*, all the βs have been integrated out.

Step 5. Generate τ^2, given μ, the $\{\beta_j\}$ and the σ_j^2, by the following accept-reject algorithm:

- Generate τ^2 from the inverse gamma distribution with shape parameter $\alpha = (J-2)/2$ and rate parameter $\beta = \frac{1}{2}\sum_{j=1}^{J}(\beta_j - \mu)^2$.
- Draw a uniform random variable U on $(0, 1)$ and accept τ^2 if

$$U < \frac{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2, \tau^2)}}{\sqrt{\sum_{j=1}^{J} 1/v_j^2(\sigma_j^2, 0)}} ;$$

else discard τ^2 and draw a new τ^2, repeating as necessary until a τ^2 is accepted.

This algorithm follows from noting that the likelihood for τ^2, given all the other parameters, is proportional to the given inverse gamma distribution. The posterior distribution of τ^2, given all the other parameters, is then proportional to this likelihood times the prior; a sample is then drawn from this posterior using accept/reject with the likelihood as the proposal distribution.

To *view* samples from the posterior and assess that the MCMC algorithm is behaving properly (*Mengersen et al.*, 1999), we consider histograms and trace plots, respectively. Trace plots illustrate the entire sequence of samples from the posterior distribution, or chain, (after the first few thousand are discarded) with the value of the random variable plotted on the vertical axis vs. the sequence index. The reader unfamiliar with MCMC sampling should note that a well-mixing algorithm should not get *stuck* at one value for many samples, should not have too many vertical outliers, and should not have a discernible periodic envelope. Note, the samples (and trace plots) have been *thinned* keeping every fifth sample from the MCMC sequence.

Of particular interest are slope parameters for each volcano, β_j, illustrated for the frictional model in Figure 5. Histograms of slope parameter samples for each volcano give reassurance that we are sampling around a common slope, near -0.2. Spread in each individual histogram reflects the uncertainty of the slope parameter for each volcano. Of course, wider histograms indicate more uncertainty.

Samples for *any* of the unknown parameters described by the posterior distribution can be visualized in this manner. For example, one might be interested in estimating the inferential variance parameters for the two flow categorizations, channelized vs. unchannelized. Descriptive illustrations of these samples are presented in Figure 6. The unknown parameters of particular interest are always dependent on the scientific questions at hand for a given problem.

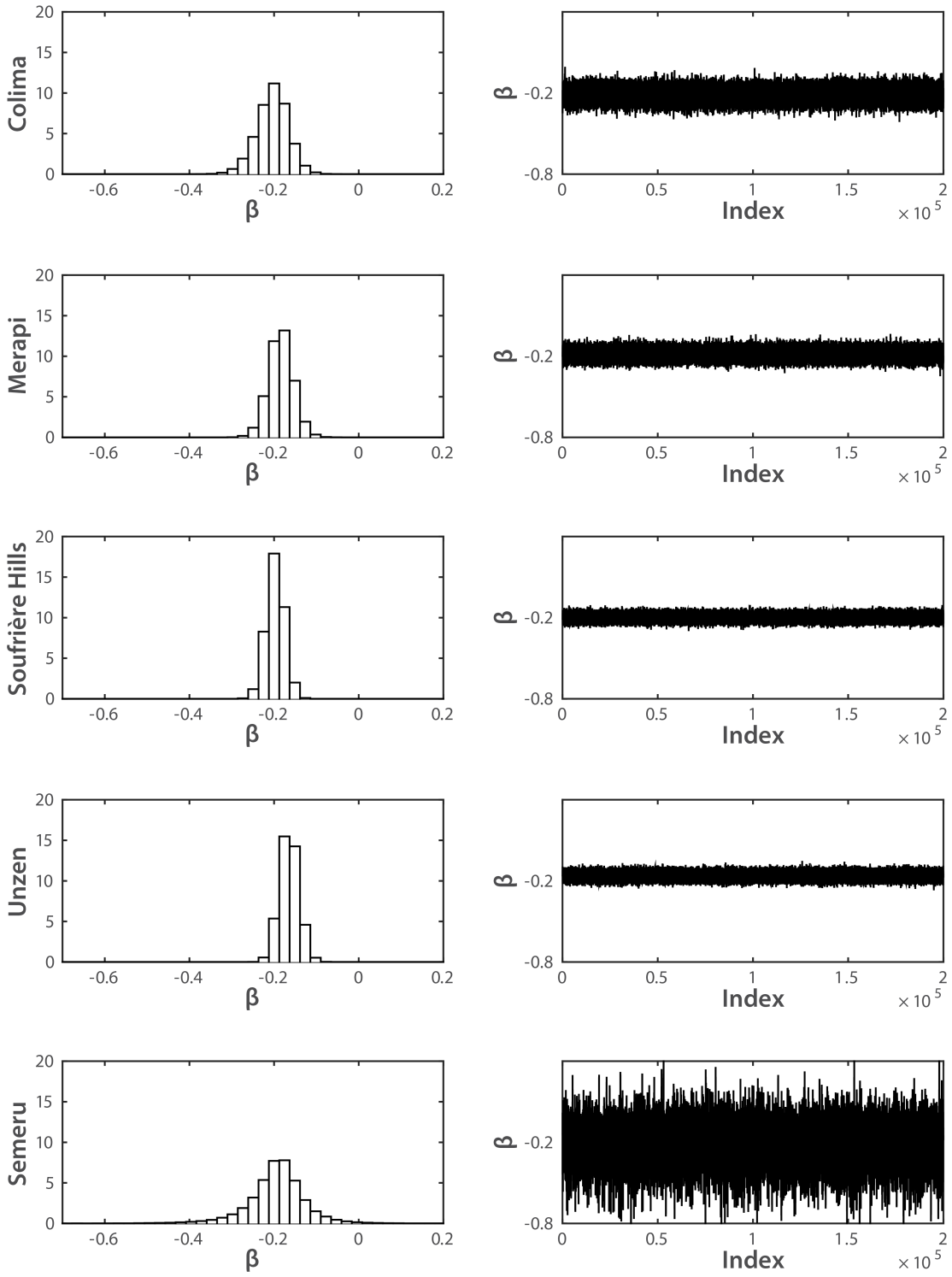

Figure 5: Left: Normalized histograms of sampled slopes for the frictional model for each of the five volcanoes considered. Right: corresponding trace plots from MCMC samples.

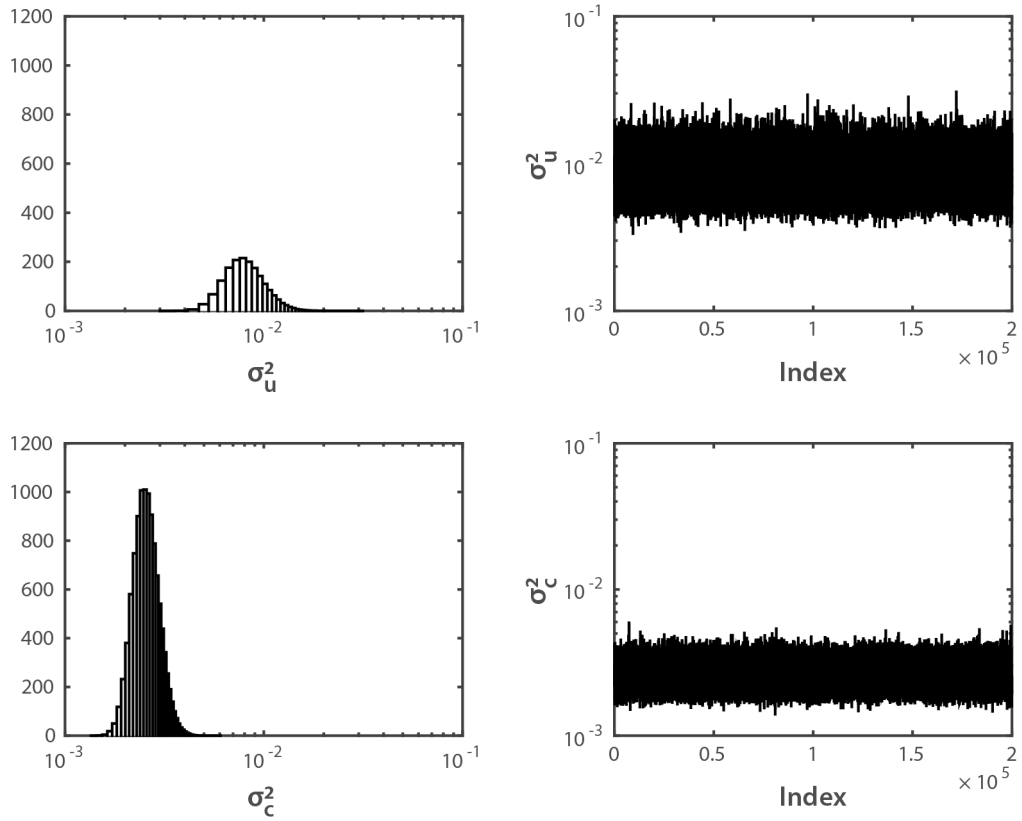

Figure 6: Left: Normalized histograms of the inferential variances, σ_u^2 (unchannelized, top) and σ_c^2 (channelized, bottom), for linear regression model applied to the frictional relationship, plotted on a log scale. Right: corresponding trace plots from MCMC samples.